

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-322088

(P 2 0 0 0 - 3 2 2 0 8 8 A)

(43) 公開日 平成12年11月24日 (2000. 11. 24)

(51) Int. Cl. <sup>7</sup>	識別記号	F I	テーマコード (参考)
G10L 15/28		G10L 3/00 511	5B075
G06F 3/16	320	G06F 3/16 320	H 5D015
17/30		15/40 370	J 9A001
G10L 15/06		15/403 310	Z
		G10L 3/00 521	A

審査請求 未請求 請求項の数 9 O L (全15頁) 最終頁に続く

(21) 出願番号 特願平11-133659

(22) 出願日 平成11年5月14日 (1999. 5. 14)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 脇坂 新路

東京都小平市上水本町五丁目20番1号 株式会社日立製作所半導体事業部内

(72) 発明者 近藤 和夫

東京都小平市上水本町五丁目20番1号 株式会社日立製作所半導体事業部内

(74) 代理人 100095913

弁理士 沼形 義彰 (外1名)

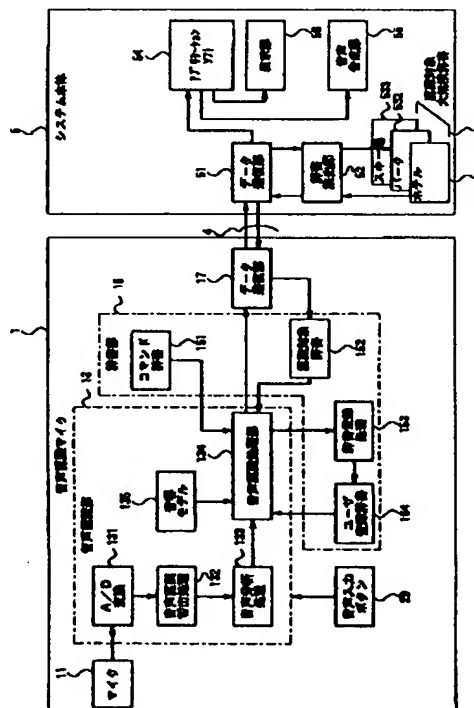
最終頁に続く

(54) 【発明の名称】 音声認識マイクおよび音声認識システムならびに音声認識方法

(57) 【要約】

【課題】 カーナビゲーションシステム、小型情報機器、ゲームなどに用いられる音声認識システムにおいて、音声認識を使い易いインターフェースにする。

【解決手段】 音声認識処理を実行する機能を有する音声認識マイク1と、システム本体5とを通信手段4で接続し、音声認識部13で認識した認識結果をシステム本体5に転送してシステムを動作させる。音声認識マイク1は、音響モデル135と音声認識処理部134とを有する音声認識部13と、コマンド辞書131と認識対象辞書152とユーザ登録辞書153と、データ通信部17とを有し、システム本体5から転送された認識対象辞書152から必要な単語をユーザ登録辞書154に登録し、通常は、コマンド辞書151とユーザ登録辞書154とを用いて音声認識を行う。



## 【特許請求の範囲】

【請求項 1】 音声認識の対象となる単語や文章を集めて辞書として定義し、音声認識結果に基づいてそれらの単語や文章を取り出して、後続する情報処理用データや文字列表示や単語が示す画像などとして出力したり、認識結果を音声合成を用いて音声として出力する音声認識システムにける音声認識機能を備えた音声認識マイクにおいて、マイク単体の機能を有するマイク部と、マイク部からのアナログ信号をデジタル信号に変換する A/D 変換器と、音声区間を検出する音声区間検出処理と、取り込んだ音声に対して音声分析する音声分析処理と、音声の特徴を音素単位でもつ音響モデルと、あらかじめ登録された辞書と音響モデルを連結して、音響モデルと連結された全ての辞書において、入力された音声の音声分析結果と照合し、確からしい認識結果を出力する音声認識処理部を備え、マイクから音声認識結果を出力することを特徴とする音声認識マイク。

【請求項 2】 音声の入力から音声認識結果を出力するまでの一連の音声認識処理を行う音声認識部と、音声認識結果を用いて新たな処理を実行するシステム本体へ認識結果を転送しシステム本体から認識対象データを転送するデータ通信部と、認識対象となる辞書を含む辞書部を有していることを特徴とする請求項 1 に記載の音声認識マイク。

【請求項 3】 データ通信部は、音声認識マイクとシステム本体を有線または無線もしくは赤外線通信で接続するインタフェースであり、通信されるデータはデジタルデータであり、その内容は、音声認識結果を表すテキストデータおよび／またはデジタル化された音声波形データおよび／または音声でない雑音 N と音声 S のレベルを示す S/N 比のデータであることを特徴とする請求項 2 に記載の音声認識マイク。

【請求項 4】 これから音声を入力することを音声認識マイクに知らせるための音声入力通知手段を接続するためのインタフェースを有することを特徴とする請求項 1 ないし請求項 3 のいずれか 1 項に記載の音声認識マイク。

【請求項 5】 あらかじめ音声認識マイクに登録したコマンド辞書と、システム本体の記憶媒体に登録された大規模な辞書から認識対象となる辞書を読み出した認識対象辞書と、ユーザの登録処理によって作成されるユーザ登録辞書とを有することを特徴とする請求項 1 ないし請求項 4 のいずれか 1 項に記載の音声認識マイク。

【請求項 6】 音声認識機能を有し入力された音声を認識した結果を出力する音声認識マイクと、音声認識結果を用いて後続する処理を実行する情報処理手段からなる音声認識システム。

【請求項 7】 上記音声認識マイクが音声認識処理部と辞書部とデータ通信部を有し、上記情報処理手段が音声認識に使用する認識対象辞書とデータ通信部を有し、前

記音声認識マイクでの音声認識結果に基づいて音声認識マイクの辞書部に前記情報処理手段の認識対象辞書の一部を転送し、転送された認識対象辞書を用いて音声認識することを特徴とする請求項 6 に記載の音声認識システム。

【請求項 8】 上記音声認識マイクの辞書部にユーザ登録辞書を設け、該ユーザ辞書に認識対象辞書の中のユーザが必要とする単語を登録し、通常の音声認識では、コマンド辞書とユーザ登録辞書を認識対象として音声認識することを特徴とする請求項 6 または請求 7 に記載の音声認識システム。

【請求項 9】 音声認識処理部と辞書部とデータ通信部を有する音声認識マイクと、音声認識結果に基づいて処理を行う情報処理手段とからなる音声認識システムの音声認識方法において、音声認識マイクの辞書部に、コマンド辞書と、情報処理手段から転送された認識対象辞書と、ユーザ登録辞書とを備え、認識対象辞書の中から最終的にユーザが必要とする複数の単語を集めてユーザ登録辞書を作成し、通常の音声認識ではコマンド辞書とユーザ登録辞書を認識対象として音声認識することを特徴とする音声認識システムの音声認識方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、音声認識システムおよび方法にかかわり、カーナビゲーションシステム、車載用 PC、カーエレクトロニクスや、PDA、ハンドヘルド PC に代表される小型情報機器、携帯型音声翻訳機、ならびに、ゲーム機器、家電機器に用いる音声認識システムであって、特に、カーナビゲーションシステムや車載用 PC、カーエレクトロニクスに代表されるカーマルチメディア分野において、認識応答時間、認識率向上の面で、使い勝手の良い音声認識システムおよび方法に関する。

## 【0002】

【従来の技術】近年、音声認識技術を用いた小型情報システムが普及しつつある。カーナビゲーションシステムをはじめとして、PDA に代表される小型情報機器、携帯型翻訳機等である。このような音声認識システムの例として、特開平 5-35776 号公報には「言語自動選択機能付翻訳装置」として、マイクから入力した操作者の音声を認識して、翻訳し、翻訳した言語の音声を出力するようにした携帯用の翻訳装置に関する技術が開示されている。

【0003】以下、図 6 を用いて、このような従来技術にかかわる音声翻訳装置の概要を説明する。図 6 は、従来技術にかかわる音声翻訳装置の構成を示すブロック図である。音声認識手段を備えた音声翻訳装置 7 は、マイク 71 と、音声認識部 72 と、翻訳部 73 と、制御部 74 と、これら各部 72、73、74 管でデータを転送するバス 75 とを有して構成される。

【0004】音声認識部72は、音声区間切出部721と、音声認識処理部722と、音声モデル格納部723と、音声認識辞書部724とを有して構成される。

【0005】翻訳部73は、翻訳語データ用メモリカード731と、音声合成部732と、表示部735とを有している。さらに、音声合成部732には、スピーカアンプ733と、スピーカ734が接続されている。

【0006】マイク71は、ユーザの音声などを電気信号に変換して入力する。

【0007】音声区間切出部721は、マイク71から入力された音声と雑音を含んだ音声信号をデジタル信号に変換するとともに音声区間を切り出し、音声区間の信号を音声認識処理部722に送る。

【0008】音声認識処理部722は、キーボード又はスイッチ等による操作信号79を受けた制御部74の指示により、マイク71、音声区間切出部721を経て、切り出された音声を音響モデル格納部723に格納された音響モデルを用いて分析する。さらに、音声認識処理部722は、分析した結果を、音声認識辞書部724に格納された標準音声パターンと比較することによって、音声認識を行う。

【0009】音響モデル格納部723には、音声認識に用いる切り出された音声区間の音響モデルが格納されている。

【0010】音声認識辞書部724は、RAM等からなり、操作者の発声に応じた標準音声パターンを格納している。この標準音声パターンは、操作者があらかじめ格納しておく。

【0011】一方、翻訳部73の翻訳語データ用メモリカード731は、ROMカード等からなり、音声認識した単語に対応する翻訳語が格納されており、翻訳語を音声合成して出力する場合には、音声データを格納している。また、この翻訳語データ用メモリカード731から、翻訳語に対応したキャラクターコードを読み込み、表示部735に表示する。翻訳語データ用メモリカード731を他の言語のものと交換することによって、音声認識した単語を複数の言語に対応して翻訳することが可能となる。

【0012】音声合成部732は、音声認識処理部722により認識された音声に対応した翻訳語を、翻訳語データ用メモリカード731から読み込み、音声信号に変換してスピーカアンプ733、スピーカ734を経て出力する。

【0013】表示部735は、翻訳装置の使用者への指示や翻訳語の文字による表示等をおこなう。

【0014】制御部74は、マイクロプロセッサ等からなり、音声翻訳装置7の各部を制御する。

【0015】このような音声認識、音声合成技術の分野は、半導体技術の向上を背景として、システムがより人間的なユーザインタフェースを提供すべきであるという

要望から、その発展が期待されている。上記従来の音声認識技術を用いた小型情報システムにおいても、カーナビゲーションシステムをはじめとして、PDAに代表される携帯型情報機器、携帯型翻訳機、さらに、音声インタフェースを持った情報家電として、今後ますます普及して行くことが予想される。

【0016】そこで、このような音声認識技術を使用した分野での実用化における課題は、認識率の向上と認識応答時間の短縮にある。従来の技術では、認識率や認識応答時間の性能を低下させないためには、認識する語数に制約を設ける必要がある。その制約の中で、あらかじめ登録しておいた単語、文に対して、その文字列が持つ統計的な話者の音声の特徴と、実際に話者が発声した音声の特徴とを比較し、確率的に一番近い値を認識結果としている。特に、雑音環境下におけるあるレベル以上の認識率を確保するには、この手法が必要不可欠である。

【0017】今後、音声認識における技術革新や、それを実現するソフトウェア、ハードウェアの性能向上により、認識する語数に制約を設けなくとも、認識率や認識応答時間の性能は向上することが考えられる。しかしながら、音声認識システムの実用的な観点から、処理量はできるだけ小さい方が認識率や認識応答時間における音声認識の単体性能、並びに音声認識を組み込んだシステム全体の性能と使い勝手の面では好ましい。また、音声認識を使い易いインタフェースにするための課題は、音声認識を音声を用いた単なる一つのユーザインタフェースにすることである。

【0018】そのためには、音声認識処理をシステム本体で行わず、マイク等のインタフェース側で実現することである。それにより、システム本体とマイク等の音声認識インタフェースは、既存のシステムに容易に接続可能となる。さらに、音声認識処理した結果をシステム本体へ転送することから、従来のアナログ音声信号をシステム本体へ転送してから音声認識する場合と比べて、環境からのノイズの影響を小さくすることができる。したがってシステム全体の認識性能を向上させることができる。そのためには、システム本体と、音声認識を実行するマイクなどのインタフェース部とを分離して使い勝手のよいシステムを提供する必要がある。

【0019】従来のカーナビゲーションシステムにおける音声認識システムでは、地名、交差点名、建物名、駅名、電話番号などの音声認識対象となる辞書を数十万単語用意して、辞書を階層的に分割して音声認識を階層的に実行する。目的の単語まで到達するまでに、数回の音声を発生し、かつ音声認識するまでくり返す。目的の単語が認識されると、例えば、目的地までのルート探索が行われる。

【0020】このようなシステムでは、認識対象となる辞書の語数は膨大であり、認識率や認識応答時間の性能を低下させないために、辞書の階層化および音声認識を

階層的に実行する。これでは、一つの目的単語を認識させるのに数回の音声認識を実行しなければならない。したがって、便利であるはずの音声認識によるインタフェースが逆に不便なものとなりシステム全体の使い勝手が悪くなってしまう。

【0021】また、辞書を階層的にせず、数十万単語からなる辞書で、はじめから目的の単語を発声する音声認識システムでは、将来、音声認識技術の革新で認識率が向上したとしても、特に、システムの低価格化においては、認識率や認識応答時間の面で十分な性能が得られない。

【0022】さらに、辞書を階層化し数回の音声認識を実行して、基本的な音声認識の性能が得られたとしても、認識率は100%にはならない。それは、人間が音的に類似した単語を聞き間違えるのと同じである。

【0023】例えば、カーナビゲーションシステムに適用した音声認識システムを図7を用いて説明する。図7は、従来の音声認識システムであり、音声認識処理をカーナビゲーションシステム本体側で行っているシステムにおける音声認識処理の流れを説明する図である。従来のシステムでは、音声認識に関わる一連の処理は、カーナビゲーションシステム本体が行っているため、カーナビゲーションシステムに限らず、他に優先順位の高い処理との競合が起り、CPUの負荷は大きくなる。また、もともと音声のインタフェースを持たないシステムにおいては、CPUの負荷に加えて、ハードウェアの改造が伴う。

【0024】ここでは、第1の発声「ホテル」の後に、第2の発声「△△△ホテル」を発した場合の音声認識処理を説明する。このシステムでは、対象別に複数の辞書が設けられている。第1の発声「ホテル」が入力される(S1)と、音声認識処理P1は、一連の音声認識処理を行って、認識結果「ホテル」を出力し、アプリケーションP2へ送出する(S2)。認識結果「ホテル」は、アプリケーションP2で、辞書選択処理P3を起動し(S3)、大規模認識対象辞書53内のホテル単語辞書531を選択する(S4)。

【0025】ホテル単語辞書531は、ホテル名の単語で構成された辞書であり、単語数は、5000単語であるとする。大規模認識対象辞書53として、ホテル単語辞書の他にパーク単語辞書531、スキー場単語辞書532などが保存されており、単語数は、それぞれ10000単語、4000単語であるとする。選択されたホテル単語辞書531は、認識対象辞書534として以降の音声認識処理P1に用いられる。

【0026】第2以降の発声「△△△ホテル」が入力されると(S5)、音声認識所利P1は音声認識処理を行い、認識結果「△△△ホテル」をアプリケーションP2に出力する(S6)。アプリケーションP2は、このデータを、目的地設定、ルート検索などの処理に渡す(S

7)。

【0027】このような方式では、ユーザは、以上の様な「ホテル」の入力から始まるやり取りを、目的地を設定することに行わなければならない。先に述べた通り、このような手法は、CPUの負荷と、辞書の単語数から、認識応答時間と認識率の面で、特に、システムトータルの低価格化において認識性能を劣化させることとなる。

【0028】

10 【発明が解決しようとする課題】以上の点に鑑み、本発明の第一の目的は、音声認識を実行する音声認識インタフェース部を音声認識結果を利用するシステム本体と分離して、システム本体へは認識した結果だけを転送する音声認識マイクを提供することにある。また、本発明の第二の目的は、性能の面から見ても使い勝手の良い音声認識マイクを提供することである。

20 【0029】すなわち、例えば、カーナビゲーションシステムに登録されている膨大な辞書の中で、ユーザがよく使う複数の辞書の単語数は、コマンドや目的地の地名など合計しても100単語以下であると想定する。そこで、音声認識マイクの認識対象単語は、ユーザが、システム本体側の膨大な辞書から必要な単語だけを、システム本体側から音声認識マイク側へ登録した単語に限って音声認識処理を実行する。したがって、認識率は100%に近くなり、また処理量も小さくなることから、低価格なハードウェアでも十分な性能が実現できる。

【0030】

30 【課題を解決するための手段】上記目的を達成するために、請求項1の発明は、音声認識の対象となる単語や文章を集めて辞書として定義し、音声認識結果に基づいてそれらの単語や文章を取り出して、後続する情報処理用データや文字列表示や単語が示す画像などとして出力したり、認識結果を音声合成を用いて音声として出力する音声認識システムにける音声認識機能を備えた音声認識マイクにおいて、マイク単体の機能を有するマイク部と、マイク部からのアナログ信号をデジタル信号に変換するA/D変換器と、音声区間を検出する音声区間検出処理と、取り込んだ音声に対して音声分析する音声分析処理と、音声の特徴を音素単位でもつ音響モデルと、あらかじめ登録された辞書と音響モデルを連結して、音響モデルと連結された全ての辞書において、入力された音声の音声分析結果と照合し、確からしい認識結果を出力する音声認識処理部を備えて構成し、マイクから音声認識結果を出力するようにした。

50 【0031】請求項2の発明は請求項1の音声認識マイクにおいて、音声の入力から音声認識結果を出力するまでの一連の音声認識処理を行う音声認識部と、音声認識結果を用いて新たな処理を実行するシステム本体へ認識結果を転送しシステム本体から認識対象データを転送するデータ通信部と、認識対象となる辞書を含む辞書部を

有して構成した。

【0032】請求項3の発明は、請求項2の音声認識マイクにおいて、データ通信部を、音声認識マイクとシステム本体を有線または無線もしくは赤外線通信で接続するインタフェースとして構成し、通信されるデータをデジタルデータとし、その内容を、音声認識結果を表すテキストデータおよび／またはデジタル化された音声波形データおよび／または音声でない雑音Nと音声Sのレベルを示すS/N比のデータとした。

【0033】請求項4の発明は、上記音声認識マイクにおいて、これから音声を入力することを音声認識マイクに知らせるための音声入力通知手段を接続するためのインタフェースを備えた。

【0034】請求項5の発明は、上記音声認識マイクにおいて、あらかじめ音声認識マイクに登録したコマンド辞書と、システム本体の記憶媒体に登録された大規模な辞書から認識対象となる辞書を読み出した認識対象辞書と、ユーザの登録処理によって作成されるユーザ登録辞書とを有して構成した。

【0035】請求項6の発明は、音声認識機能を有し入力された音声認識した結果を出力する音声認識マイクと、音声認識結果を用いて後続する処理を実行する情報処理手段とから音声認識システムを構築した。

【0036】請求項7の発明は、上記音声認識システムにおいて、上記音声認識マイクが音声認識処理部と辞書部とデータ通信部を有し、上記情報処理手段が音声認識に使用する認識対象辞書とデータ通信部を有し、前記音声認識マイクでの音声認識結果に基づいて音声認識マイクの辞書部に前記情報処理手段の認識対象辞書の一部分を転送し、転送された認識対象辞書を用いて音声認識するようにした。

【0037】請求項8の発明は、上記音声認識システムにおいて、上記音声認識マイクの辞書部にユーザ登録辞書を設け、該ユーザ辞書に認識対象辞書の中のユーザが必要とする単語を登録し、通常の音声認識では、コマンド辞書とユーザ登録辞書を認識対象として音声認識するようにした。

【0038】請求項9の発明は、音声認識処理部と辞書部とデータ通信部を有する音声認識マイクと、音声認識結果に基づいて処理を行う情報処理手段とからなる音声認識システムの音声認識方法において、音声認識マイクの辞書部に、コマンド辞書と、情報処理手段から転送された認識対象辞書と、ユーザ登録辞書とを備え、認識対象辞書の中から最終的にユーザが必要とする複数の単語を集めてユーザ登録辞書を作成し、通常の音声認識ではコマンド辞書とユーザ登録辞書を認識対象として音声認識するようにした。

【0039】

【発明の実施の形態】以下、本発明に係る各実施形態を、図1から図5を用いて説明する。図1は、本発明に

かかる音声認識マイクの機能構成を示すブロック図である。

【0040】図1に示す音声認識マイク1は、音声認識した結果3を出力する。この音声認識マイク1の出力3は、音声認識した結果に限らず、音声認識に関わる情報であってよい。例えば、本来のマイクの基本機能である音声や周囲の音を集音して、従来はアナログ信号として伝達していたものをデジタル化して出力したデジタル波形信号である。また、音声でない雑音Nのレベルと、音声Sのレベルを相対的な比で表したS/N比のデータである。これらのデジタル化された情報は、音声認識マイクとシステム本体を組み合わせる新たなシステムを構築する場合のアプリケーションに必要な基本情報である。

【0041】音声認識マイク1は、マイク11と、音声認識部13と、辞書部15と、データ通信部17との機能ブロックで構成される。

【0042】マイク11は、音声や雑音を取り込むもので、従来から有るコンデンサマイクなどで構成され、指向性を有している。

【0043】音声認識部13は、入力された音声や雑音から音声だけを検出して、音声分析を行う。さらに、音声認識部13は、あらかじめ登録された辞書と、音声の特徴を音素単位でもつ音響モデルから、登録された全ての辞書と音響モデルを連結して、実際に入力された音声の音声分析結果と照合して、確からしい認識結果を出力する。

【0044】辞書部15は、音声認識の対象となる辞書が格納される。辞書部15に格納される辞書には、あらかじめ音声認識マイク1に登録しておくコマンド辞書と、システム本体の記憶媒体に登録されている大規模な認識辞書から転送された認識対象辞書と、コマンド辞書や認識対象辞書からユーザが必要とする単語のみ取り出して登録して構成するユーザ登録辞書がある。辞書部15は、音声認識結果、コマンド辞書に登録された「ユーザ辞書へ登録（辞書へ登録）」を受けて、ユーザ登録辞書を作成する。

【0045】データ通信部17は、音声認識部の音声認識結果を、音声認識マイク1が接続されたシステム本体あるいは音声認識マイク1を使用しているシステム本体に、転送するための処理を行う。また、音声認識結果と同様に、デジタル化された音声信号や、音声に関わる情報を転送する。

【0046】図2を用いて、音声認識マイク1のハードウェア構成を説明する。音声認識マイク1は、マイク11と、アンプ21と、A/D変換器22と、CPU23と、ROM24と、RAM25と、有線インタフェース26-1と、赤外線インタフェース（IR）26-2と、無線インタフェース26-3と、音声認識モードインタフェース26-4と、これらを相互に接続するシステムバス27と、音声入力ボタン29とを有して構成さ

れる。

【0047】マイク11は、図1に示したマイク11と同じである。

【0048】アンプ21は、抵抗、コンデンサなどの電子部品で構成されたアンプであり、雑音を除去するためのハイパスフィルタや、バンドパスフィルタを含んでいる。

【0049】A/D変換器22は、アンプ21を経由してマイク11から入力された音声や雑音のアナログ信号をデジタル信号に変換する。A/D変換器22は、システムバス27に接続されている。

【0050】CPU23は、音声認識マイク1において、音声認識および辞書登録ならびにデータ通信の全ての処理をソフトウェアで行う中央処理ユニットあるいはCPUコアである。

【0051】ROM24には、音声認識マイクシステムの初期化および一連の音声認識および辞書登録ならびにデータ通信の全ての処理をソフトウェアで実行するためのプログラムが、書き込まれている。また、ROM24には、音声認識に必要な音響モデルや辞書、文法なども書き込まれている。

【0052】RAM25は、一連の音声認識、辞書登録、データ通信の全ての処理をソフトウェアで実行するためのプログラムをアクセスの高速なメモリに転送して実行するためのメモリであり、また、プログラム実行中に必要なワークエリアを確保するためのメモリである。さらに、電源を切ってもユーザの登録した辞書などが消えないようにするためのメモリである。

【0053】有線インタフェース26-1は、音声認識マイク1と接続されたシステム本体の間で情報(データ31)をやり取りするためのインタフェースである。システム本体とは、有線で接続され、データビット幅は、シリアルでもパラレルでもよい。

【0054】データ31は、音声認識マイク1とシステム本体の間で、有線を介して双方向に転送されるデータである。その第1は、音声認識マイク1から出力される認識結果である。認識結果は、テキストデータの文字情報であっても、コード化されたデータもよい。第2は、認識対象となる辞書データである。辞書データは、システム本体から音声認識マイク1へ転送される。第3は、音声でない雑音Nと音声Sのレベルを示すS/N比のデータである。第4は、システム本体で音声認識モードに入っていることを伝えるための情報である。例えば、システム本体側のカーナビゲーションシステムにおいて、音声認識モードとしてリモコンの発話ボタンが押された場合などである。そこで、音声認識マイク1は、入力された音声に対して、音声認識処理を実行する。

【0055】赤外線インタフェース26-2は、音声認識マイク1と接続されたシステム本体の間で情報(データ32)をやり取りするためのインタフェースである。

システム本体とは、赤外線IRを使った無線でインタフェースされ、データビット幅は、シリアルでもパラレルでもよい。

【0056】データ32は、音声認識マイク1とシステム本体の間で、赤外線通信方式を介して双方向に転送されるデータである。その内容はデータ31と同じである。

【0057】無線インタフェース26-3は、音声認識マイク1と接続されたシステム本体の間で情報(データ33)をやり取りするためのインタフェースである。システム本体とは、無線LANなどを使った無線でインタフェースされ、データビット幅は、シリアルでもパラレルでもよい。

【0058】データ33は、音声認識マイク1とシステム本体の間で、無線を介して双方向に転送されるデータである。その内容はデータ31と同様である。

【0059】ここで、システム本体とのインタフェースをとる有線インタフェース26-1、赤外線インタフェース26-2、無線インタフェース26-3は、音声認識マイク1において、どれか1つだけのインタフェースをそなえていてもよいし、全てのインタフェースを備えていてもよい。

【0060】音声認識モードインタフェース26-4は、上記システム本体で音声認識モードに入っていることを伝えるための情報において、直接に音声認識マイク1に音声認識モードに入っていることを伝えるためのインタフェースである。音声認識モードインタフェース26-4には、音声入力状態を通知する音声入力ボタン29が接続されている。直接に音声認識マイク1に音声認識モードに入っていることを伝えるための意味は、例えば、システム本体は、音声認識マイク1からの認識結果31~33を一方的に受けて、システム本体が別の処理や仕事をする場合などに有効である。このような手法が必要な場合は、音声リモコンに適応した場合が考えられる。

【0061】音声認識マイク1は、複数のLSIやICで構成してもよいし、ASIC等の一つの半導体素子上に構成してもよい。ASICであれば、CPU23はCPUコアとして構成される。

【0062】図3を用いて、音声認識マイク1と音声認識の結果を用いて動作するシステム本体5を、無線あるいは有線で接続し、音声認識をインタフェースに持つシステム構成と、処理の流れを説明する。例えば、本願におけるシステムの一例として、カーナビゲーションシステムがあげられる。

【0063】このシステムは、音声認識マイク1とシステム本体5を、無線あるいは有線4などの通信方式で接続して構成される。

【0064】音声認識マイク1は、例えば、カーナビゲーションシステムにおいては、車内のサンバイザー、シ



ートベルト、ステアリングコラム、ピラーやハンドルに設置されるか内蔵される。または、音声認識マイク1は、システムをコントロールするリモコンに内蔵される。

【0065】カーナビゲーションシステム本体5は、オーディオシステムや空調システムと一体化され、ディスプレイも含めて車内に搭載される。

【0066】音声認識マイク1は、マイク11と、音声認識部13と、辞書部15と、データ通信部17と、音声入力ボタン29とを有して構成される。

【0067】音声認識部13は、A/D変換器131と、音声区間切出部132と、音声分析処理部133と、音声認識処理部134と、音響モデル格納部135とを有して構成される。

【0068】辞書部15は、コマンド辞書151と、認識対象辞書152と、辞書登録処理部153と、ユーザ登録辞書154とを有して構成される。

【0069】データ通信部17は、音声認識処理部134によって認識された結果をシステム本体5に転送する処理を行うデータ通信部である。あるいは、システム本体5から転送されてくる認識対象の辞書データや音声認識モードに入ったことを伝えるための情報を受け取る。

【0070】システム本体5は、例えば、カーナビゲーションシステムとして構成される。この場合、システム本体5は、データ通信部51と、辞書読出部52と、認識対象大規模辞書53と、アプリケーションソフト54と、音声合成部55と、表示部56とを有して構成される。

【0071】A/D変換器131は、マイク11から出力されたアナログ信号をデジタル信号に変換して、音声区間切出部133に出力する。

【0072】音声区間切出部132は、デジタル信号に変換された音声や雑音を含む入力信号から音声を切り出して、音声分析処理部133へ出力する。

【0073】音声分析処理部133は、音声を分析して分析結果を音声認識処理部134へ出力する。

【0074】音声認識処理部134は、音声分析結果と音響モデルと辞書を用いて、確率的に音声照合の一連の処理を行い、確からしい認識結果を出力する音声認識部である。

【0075】音響モデル格納部135には、音声認識に必要な音響モデルが格納される。実用化されつつある一般的な音声認識システムでは、あらかじめ声を登録しなくても、誰が話し手でもその声を認識できるいわゆる「不特定話者対応」が主流になっている。このような音声認識に用いられる音響モデルとしては、例えば、隠れマルコフモデル(HMM:Hidden Markov Model)を用いることができる。

【0076】コマンド辞書151は、あらかじめ音声認識マイク1に登録されている辞書で、音声認識処理を

行するジャンルや、音声認識処理の命令などの単語が記述されている。

【0077】認識対象辞書152は、音声認識の対象となる辞書であり、システム本体5から必要に応じて、転送されてくる。

【0078】辞書登録処理部153は、認識対象となっている認識対象辞書152の中から、常に音声認識マイク1の中に登録しておきたい単語に対して、ユーザがコマンド辞書151の「辞書へ登録」を音声入力して、認識結果として「辞書へ登録」が音声認識処理部134から出力された場合に登録を行う処理部である。

【0079】ユーザ登録辞書154は、常に音声認識マイク1の中に登録しておきたい単語を登録する辞書である。

【0080】図4を用いて、図3に示した辞書の構成について説明する。あらかじめ音声認識マイク1に登録されたコマンド辞書151は、例えば300語の単語数からなる辞書として構成され、図4(a)に示すデータD151として例示される。この例では、コマンド辞書151は、認識対象ジャンルを表す「ホテル」、「パーク」、「スキー場」、「ゴルフ場」を始め、処理の命令を表す「辞書へ登録」、「辞書から削除」、「辞書の内容」などの単語で構成されている。

【0081】音声認識対象辞書152は、音声認識の対象となる辞書であり、システム本体5の認識対象大規模辞書53から、必要に応じて転送されてくる。音声認識マイク1は、コマンド辞書151と、この認識対象辞書152の単語の範囲に限り、入力音声に対して音声認識処理を行う。

【0082】認識対象辞書152は、5000語ぐらいの単語数からなり、コマンド辞書のジャンルを指定するコマンドに対応した単語からなる辞書として構成される。例えば、ジャンルを指定するコマンド「ホテル」に対応して、図4(b)のデータD152として例示するように、「〇〇〇ホテル」、「△△△ホテル」、「ホテル□□□□」、「◇◇◇◇旅館」…「××××ホテル」、「☆☆☆ホテル」などのホテル名を表す単語のみが登録されている。

【0083】ユーザ登録辞書データD154は、ユーザ登録辞書154の具体的な内容の一例を示している。例えば、ユーザが登録しておきたい単語で構成されており、ユーザ登録辞書154を構成している単語の数は、100単語ぐらいとする。

【0084】ユーザ登録辞書154は、100語程度の単語数からなる辞書として構成される。ユーザ登録辞書154は、ユーザが登録しておきたい単語を、コマンド辞書151や認識対象辞書152から取り出して登録した辞書として構成され、図4(c)のデータD154として例示した、「△△△ホテル」、「東京〇〇〇ランド」、…「自宅へ戻る」など、場所や処理の命令などの

単語が登録されている。

【0085】つぎに、システム本体5側の処理について説明する。システム本体5のデータ通信部51は、音声認識マイク1によって認識された結果をシステム本体5で受け取るためのデータ通信部である。あるいは、システム本体5から音声認識マイク1に対して、認識対象の辞書データや音声認識モードに入ったことを伝えるための情報を転送する。

【0086】辞書読出部52は、CD-ROMやDVDに代表される大記憶容量の記憶媒体に保存されている認識対象大規模辞書53の中から、認識対象の辞書として分類されている辞書を読み出し、データ通信部51から、音声認識マイク1へ転送する。

【0087】認識対象大規模辞書53は、認識対象の辞書が各項目ごとに分類されてCD-ROMやDVDに保存されている。例えば、ホテル単語辞書531は、ホテル名の単語で構成された辞書であり、単語数は、5000単語であるとする。認識対象辞書として、ホテル単語辞書の他にパーク単語辞書532、スキー場単語辞書533などの各種の認識対象ジャンルが保存されており、単語数は、それぞれ10000単語、4000単語であるとする。

【0088】アプリケーションソフト54は、カーナビゲーションシステムの主な処理を行うアプリケーションソフトであり、GPSの処理やナビゲーションの処理や音声インタフェースなどのシステム全般にわたる処理を行う。

【0089】表示部56は、液晶表示装置(LCD)などから構成され、地図や進行状況、音声認識結果に対する情報などを表示する。

【0090】音声合成部55は、進行状況や音声認識結果のコールバック、音声案内等を音声合成して処理する。

【0091】図5を用いて、音声認識マイク1とシステム本体5内での音声認識処理を説明する。本発明の音声認識システムは、図7に示した従来の音声認識システムに対して、音声認識処理を音声認識マイク1側で行うので、カーナビゲーションシステムに限らず、システム本体5での処理は既存のままでよく、システム側のCPUの負荷は変わらない。また、もともと音声のインタフェースを持たないシステムにおいても、システム本体のCPUの負荷は変わらない上、ハードウェアの改造も小さな変更ですみ、通信インタフェースを持っているシステムにおいては、音声インタフェースを持たない既存のハードウェアで、音声による操作を実現できる。

【0092】図5において、破線から上は音声認識マイク1における辞書の登録処理と音声認識処理とこの処理に用いる辞書を、破線から下はシステム本体5における音声認識結果の利用と辞書を示している。音声認識マイク1側は、一連の音声認識を行う。この処理には、辞書

登録前に実施するユーザ登録辞書作成処理と、ユーザ登録辞書を用いた音声認識処理を行う。

【0093】まず、登録前に第1の発声「ホテル」が入力される(S11)と、コマンド辞書151を用いて音声認識処理P11を実行し、音声認識結果「ホテル」を出力する(S12)。認識結果「ホテル」は、システム本体5側へ転送され、システム本体側のアプリケーション54で、辞書選択処理P52を起動し(S13)、認識対象を網羅した認識対象大規模辞書53の内のホテル単語辞書531を選択する。

【0094】選択されたホテル単語辞書531は、システム本体5から、音声認識マイク1側へ転送され、認識対象辞書152に格納される(S14)。

【0095】第2以降の発声「△△△ホテル」が入力されると(S15)、認識対象辞書152に格納されたホテル辞書データD531を用いて音声認識処理P11を行い、認識結果「△△△ホテル」を出力する(S16)。認識結果「△△△ホテル」は、システム本体側へ転送され、システム本体5のアプリケーション54で、目的地設定、ルート検索などの処理P57に渡される(S17)。

【0096】第3の発声「辞書へ登録」が入力される(S18)と、音声認識処理P11は、コマンド辞書151を用いて音声認識処理を行い認識結果「辞書へ登録」を出力する(S19)。認識結果「辞書へ登録」を受けて、辞書登録処理153は、音声認識マイク側のユーザ登録辞書154に、「△△△ホテル」を登録する(S20)。ユーザ登録辞書154の単語数は、100単語とする。

【0097】登録後は、ユーザは、目的地を第1の発声として、いきなり「△△△ホテル」を発声する(S21)と、ユーザ登録辞書152に格納されたユーザ登録辞書データD154を用いて音声認識処理P11を行い、認識結果「△△△ホテル」を出力する(S16)。認識結果「△△△ホテル」は、システム本体側へ転送され、システム本体5のアプリケーション54で、目的地設定、ルート検索などの処理P57に渡される(S17)。

【0098】このように、本発明によれば、必要な単語を選択してユーザ辞書154に登録した後は、ユーザは直ちに「△△△ホテル」と発声するだけで音声認識が実行される。さらに、先にも述べた通り、ユーザ登録辞書154の単語数は、100単語と少ないので、認識応答時間および認識率の面で、良好な認識性能を発揮することができ、音声認識インタフェースの向上が期待できる。

【0099】さらに、この音声認識システムは、音声認識処理におけるシステム本体の負荷が大幅に軽減されるので、音声認識マイクからの認識結果を受信できるインタフェースを追加するだけで、既存の様々なシステムに



応用できる。

【0100】

【発明の効果】本発明によれば、カーナビゲーションシステム、小型情報システム、ゲームに用いられる音声認識システムにおいて、実際に使用する環境で、雑音のレベルに合わせて音声区間検出用しきい値の設定を自動化し、自動しきい値設定による音声区間検出および、認識性能が実環境下でも劣化しない、良好な音声認識システムを実現することができる。

【図面の簡単な説明】

【図1】本発明にかかる音声認識マイクの機能の概要を説明するブロック図。

【図2】本発明にかかる音声認識マイクのハードウェア構成を示すブロック図。

【図3】本発明にかかる音声認識システムの概要を説明するブロック図。

【図4】本発明にかかる音声認識システムにおける辞書構成を説明する図。

【図5】本発明にかかる音声認識システムの動作を説明する図。

【図6】従来の音声認識システムを使用した携帯型翻訳装置の構成を説明するブロック図。

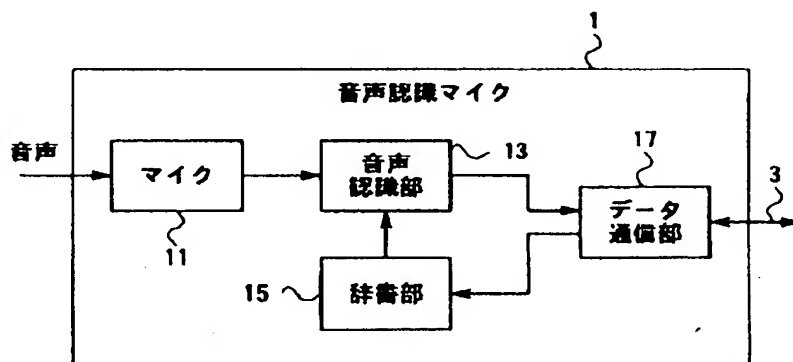
【図7】従来の音声認識システムの動作を説明する図。

【符号の説明】

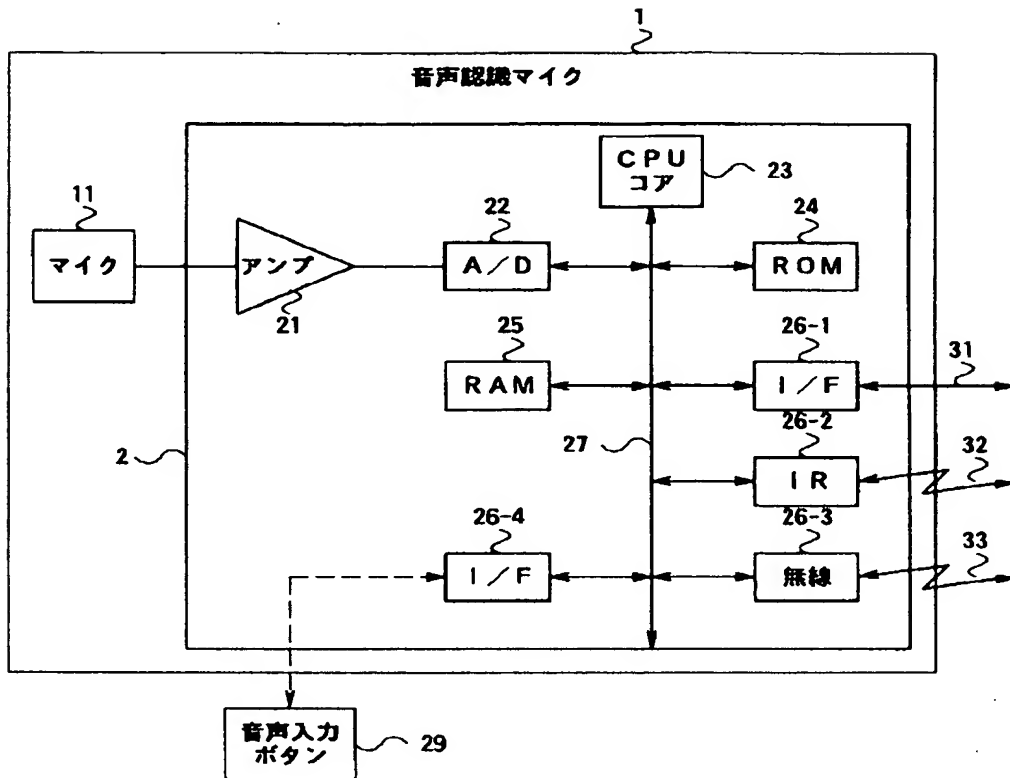
- 1 音声認識マイク
- 11 音声認識マイク
- 13 音声認識部
- 131 A/D変換器
- 132 音声区間検出部

- 133 音声分析処理部
- 134 音声認識処理部
- 135 音響モデル格納部
- 15 辞書部
- 151 コマンド辞書
- 152 認識対象辞書
- 153 辞書登録処理部
- 154 ユーザ登録辞書
- 17 データ通信部
- 21 アンプ
- 22 A/D変換器
- 23 CPU
- 24 ROM
- 25 RAM
- 26 インタフェース
- 29 音声入力ボタン
- 3 認識結果
- 4 通信手段
- 5 システム本体
- 51 データ通信部
- 52 辞書読出部
- 53 認識対象大規模辞書
- 531 ホテル単語辞書
- 532 パーク単語辞書
- 533 スキー場単語辞書
- 54 アプリケーションソフ
- 55 音声合成部
- 56 表示部

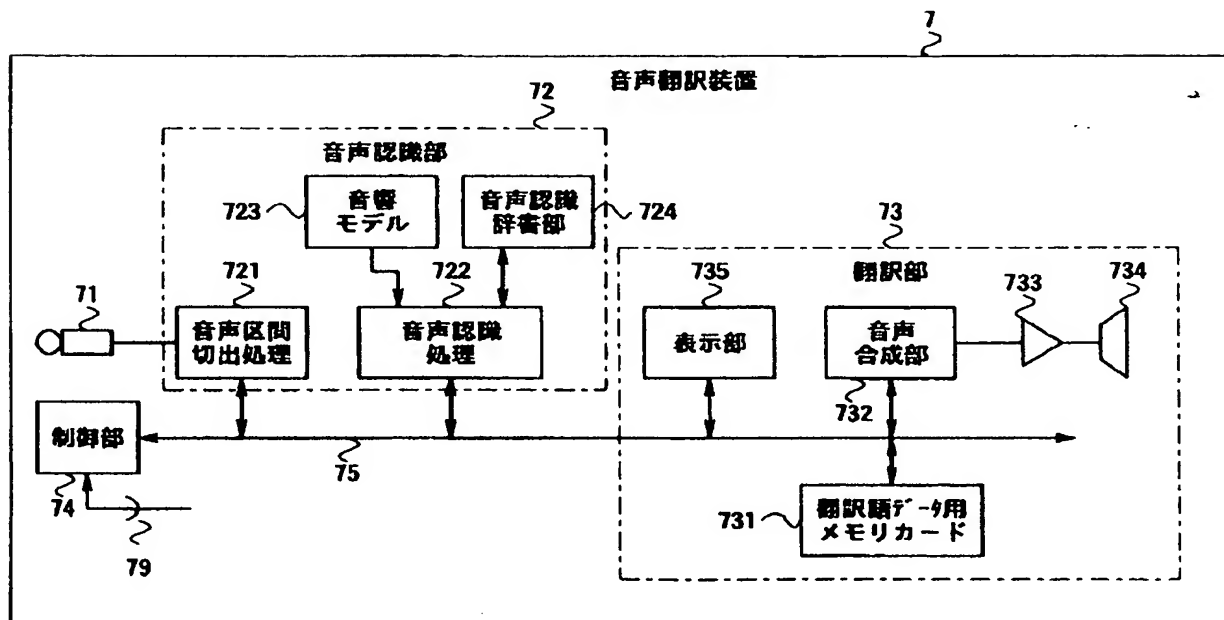
【図1】



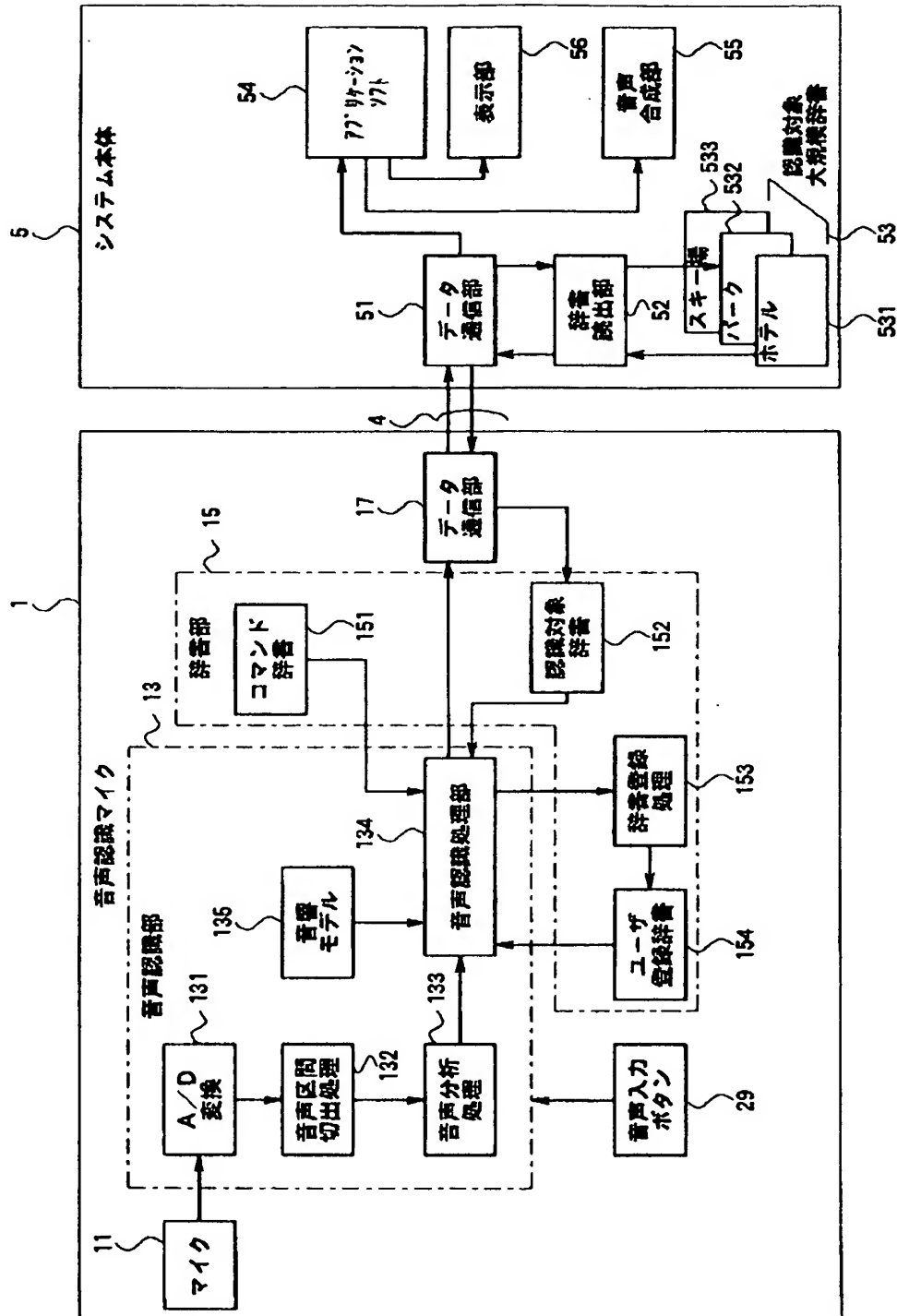
【図2】



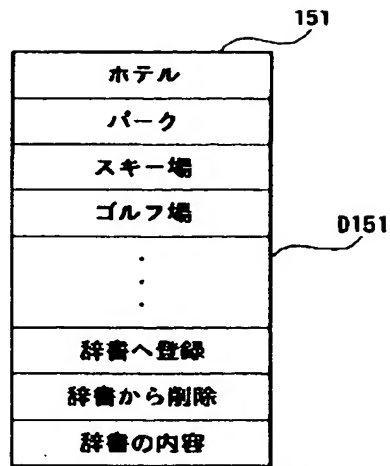
【図6】



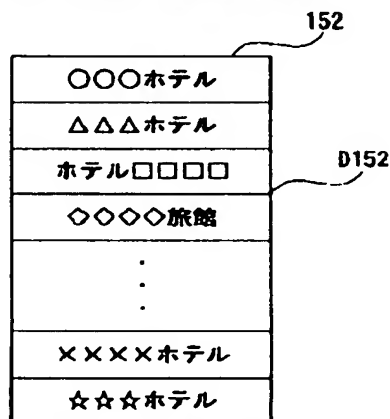
【図 3】



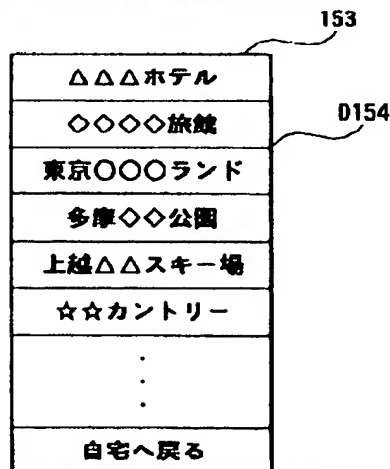
【図 4】



(a) コマンド辞書 (単語数300語)

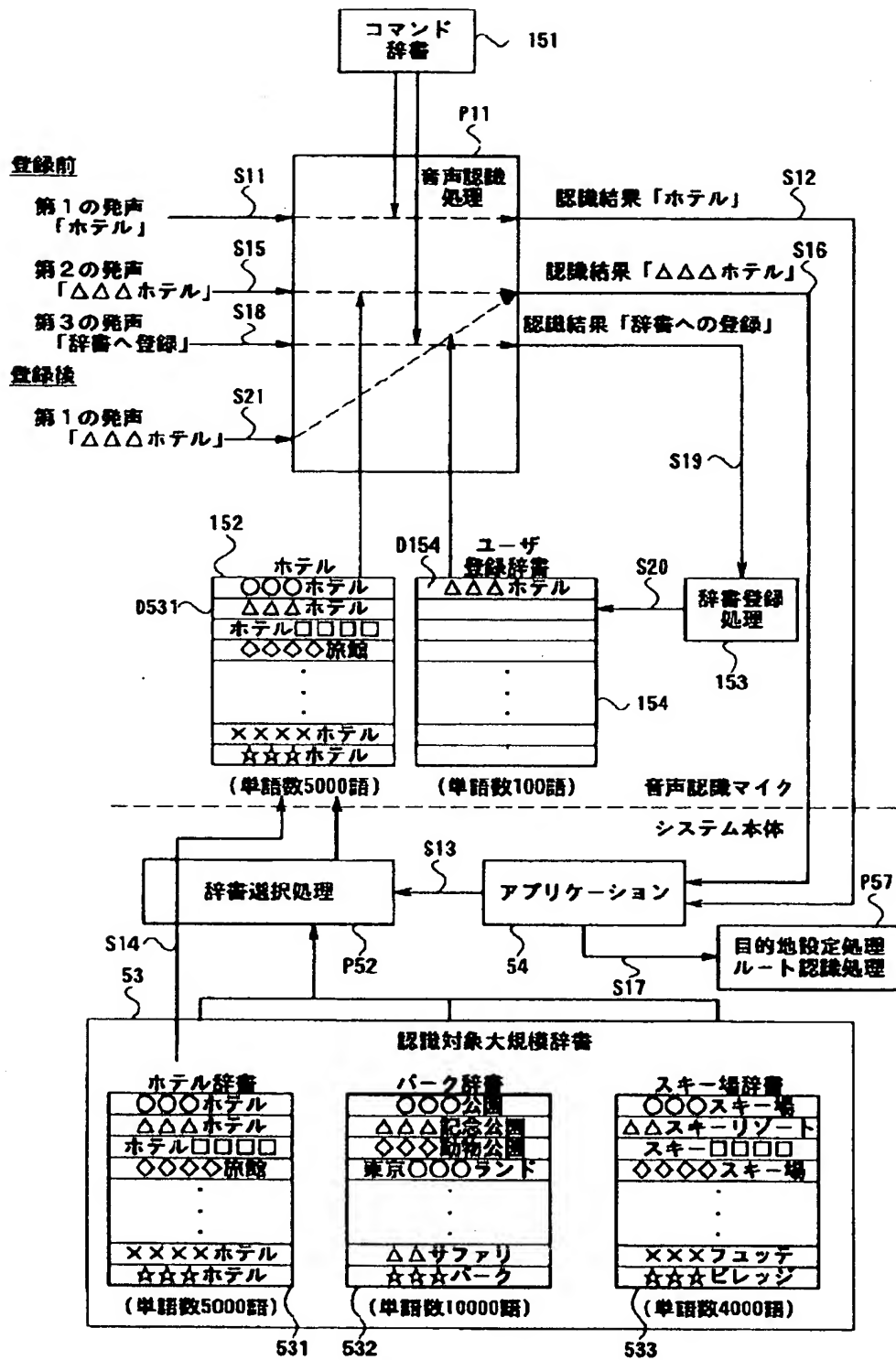


(b) 認識対象辞書 (単語数500語)

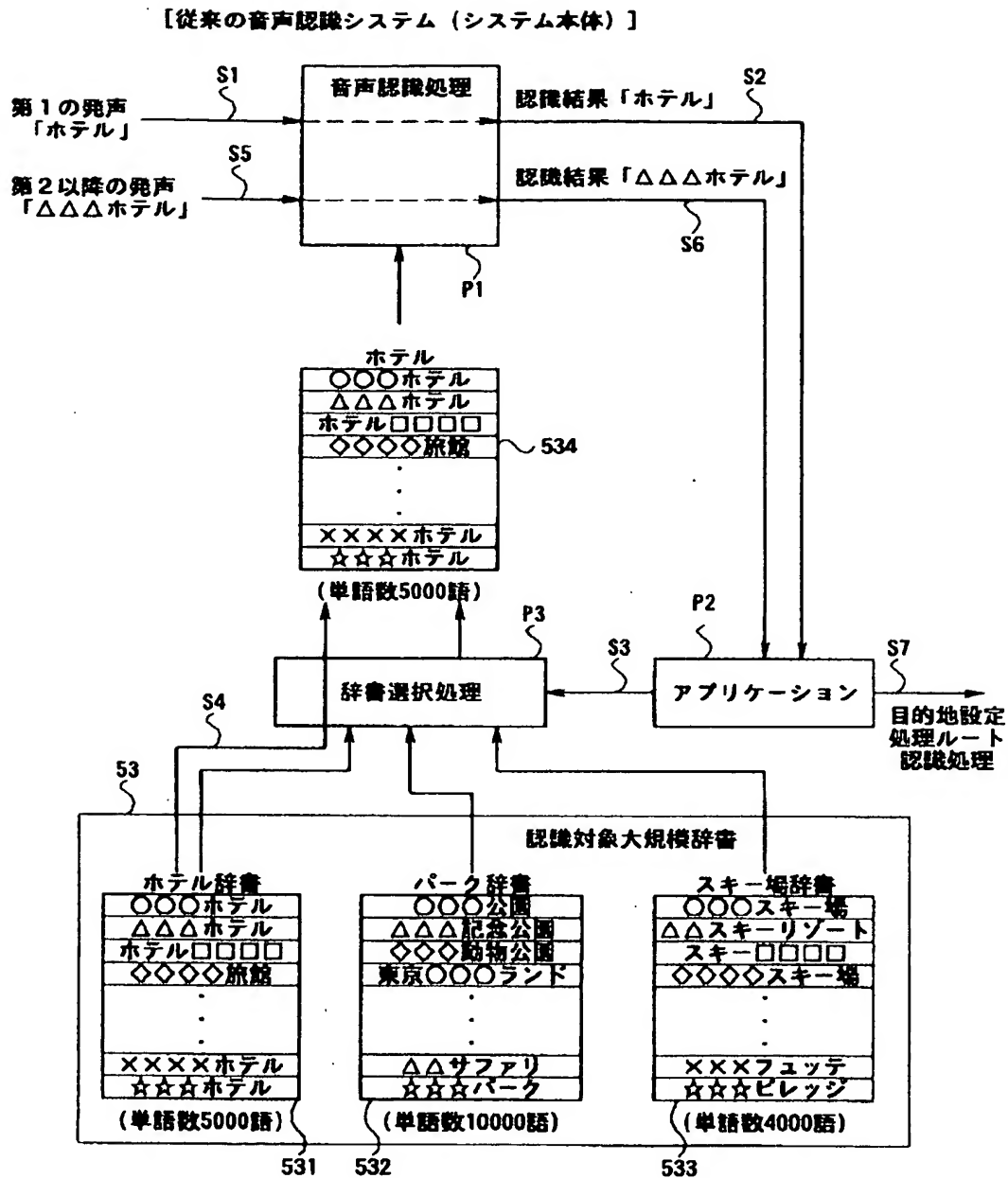


(c) ユーザ登録辞書 (単語数100語)

【図5】



【図 7】



フロントページの続き

(51) Int. Cl.<sup>7</sup>

識別記号

F I

ターマコード (参考)

G 1 0 L 3/00

5 7 1 A

(72) 発明者 小窪 浩明  
東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内

(72) 発明者 畑岡 信夫  
東京都国分寺市東恋ヶ窪 1 丁目 280 番地  
株式会社日立製作所中央研究所内



F ターム(参考) 5B075 ND02 PP07 PP22 PQ02 PQ04  
PQ05 UU01  
5D015 DD02 GG03 LL09 LL11  
9A001 CC05 EE05 HH16 HH17 JJ77  
JZ76

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-322088

(43)Date of publication of application : 24.11.2000

(51)Int.Cl.

G10L 15/28

G06F 3/16

G06F 17/30

G10L 15/06

(21)Application number : 11-133659

(71)Applicant : HITACHI LTD

(22)Date of filing : 14.05.1999

(72)Inventor : WAKIZAKA SHINJI

KONDO KAZUO

KOKUBO HIROAKI

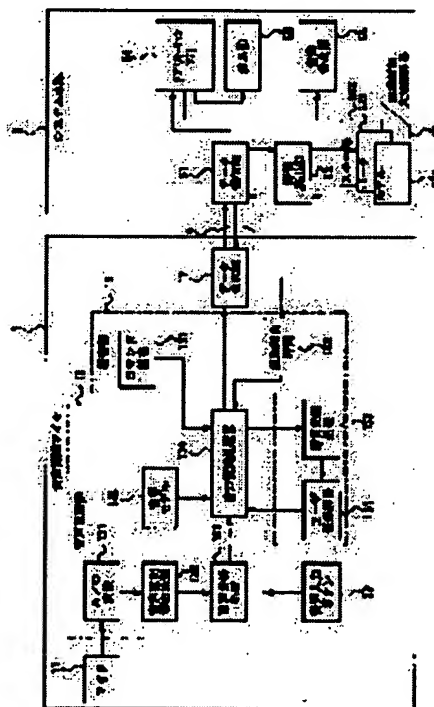
HATAOKA NOBUO

## (54) SPEECH RECOGNITION MICROPHONE AND SPEECH RECOGNITION SYSTEM AND SPEECH RECOGNITION METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide an easy-to-use speech recognition interface in regard to a speech recognition system used in car navigation systems, compact information appliances, games, or the like.

SOLUTION: A speech recognition microphone 1 having a function to execute a speech recognition process and a system main body 5 are connected by a communication means 4, and a recognition result recognized by a speech recognition part 13 is transferred to the system main body 5 for operating a system. The speech recognition microphone 1 has the speech recognition part 13 having an acoustic model 135 and a speech recognition processing part 134, a command dictionary 151, a recognition target dictionary 152, a user registration dictionary 154, and a data communication part 17, registers necessary words from the recognition target dictionary 152 transferred from the system main body 5 to the user registration dictionary 154, and performs speech recognition normally using the command dictionary 151 and the user registration dictionary 154.



## LEGAL STATUS

[Date of request for examination]

30.07.2003

[Date of sending the examiner's decision of

BEST AVAILABLE COPY

rejection]

[Kind of final disposal of application other than  
the examiner's decision of rejection or  
application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office